PENNSTATE ARL

Applications of Data Mining in Automated ISHM and Control for Complex Engineering Systems

Karl M. Reichard
Applied Research Laboratory
Penn State University
Voice: 814-863-7681
Email: kmr5@psu.edu

PENNSTATE ARL

Data Mining

- **Data mining**, or knowledge discovery, is the computer-assisted process of digging through and analyzing enormous sets of data and then extracting the meaning of the data.
- Data mining tools are used to identify correlations and to predict behaviors and future trends from data.
- Businesses use data mining to search and analyze databases for hidden patterns and may find trends and predictive information that experts may miss because it lies outside their expectations or because they never thought to look for correlations there.

PENNSTATE ARL

Data Mining

- Data mining derives its name from the similarities between searching for valuable information in a large database and mining a mountain for valuable minerals.
- For businesses, the economics of data mining have many analogies to traditional mineral mining – pay for access to public data and make money from mining it.
- A **data warehouse** is a repository where data located in disparate databases are consolidated. Data warehouses store large quantities of data by specific categories so users can easily retrieve, interpret and sort the contents.

PENNSTATE ARL

Elements of Data Mining

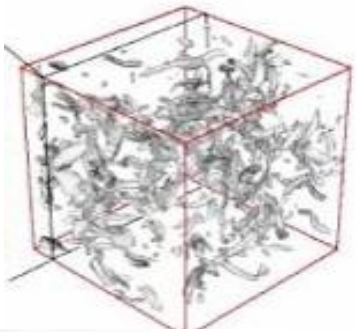
- Data preparation
- Databases and data warehouses
- Machine learning
  - Statistical models
  - Clustering
- Data Models
- Knowledge Representation
  - Association Rules
  - Classification Rules
  - Decision trees
- Data transformation

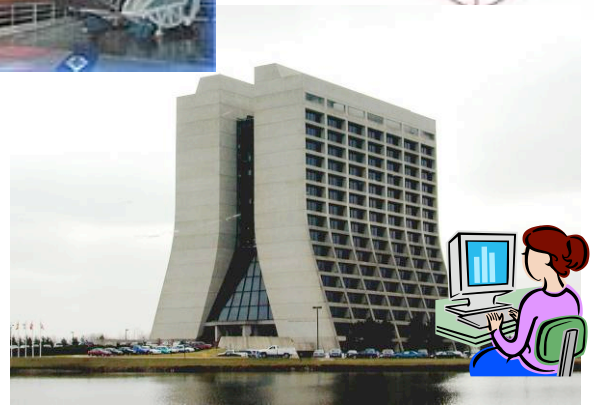
PENNSTATE ARL

Data Mining as Scientific Evolution

- **Empirical Science**
  - Scientist gathers data by direct observation
  - Scientist analyzes data
- **Analytical Science**
  - Scientist builds analytical model
  - Makes predictions.
- **Computational Science**
  - Simulate analytical model
  - Validate model and makes predictions
- **Science - Informatics**
  - Data captured by instruments
    Or data generated by simulator
  - Processed by software
  - Placed in a database / files
  - Scientist analyzes database / files

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - \mathrm{K}\frac{c^2}{a^2}$$



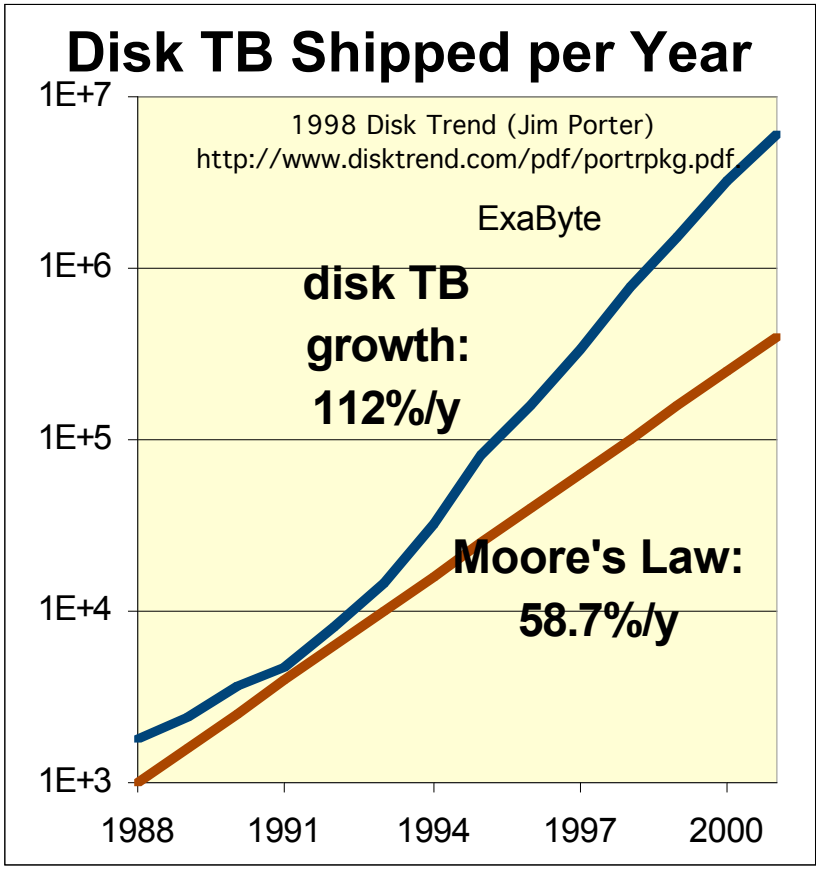


Jim Gray, "Online Science the New Computational Science," Microsoft Research.

PENNSTATE ARL

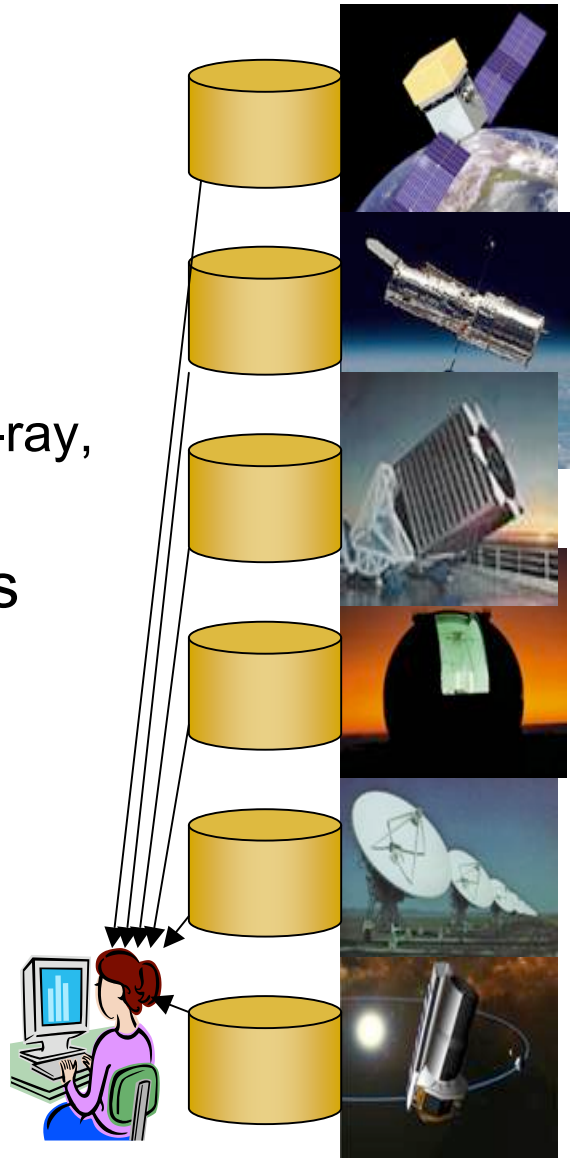
Leveraging Technology Trends

- Moore's Law
  - Performance/Price doubles every 18 months
  - 100x per decade
- Computer storage capacity is actually beating Moore's law
- Metcalfe's law
  - Utility of computer networks grows as the number of possible connections: $O(N^2)$



Example: World Wide Telescope Virtual Observatory

- Premise: Most data is (or could be online) so, the Internet is the world's best telescope:
  - It has data on every part of the sky, in every measured spectral band: optical, x-ray, radio..
  - It is up when you are up. The "seeing" is always great (no working at night, no clouds no moons no..).
  - It's a smart telescope: links objects and data to literature on them.



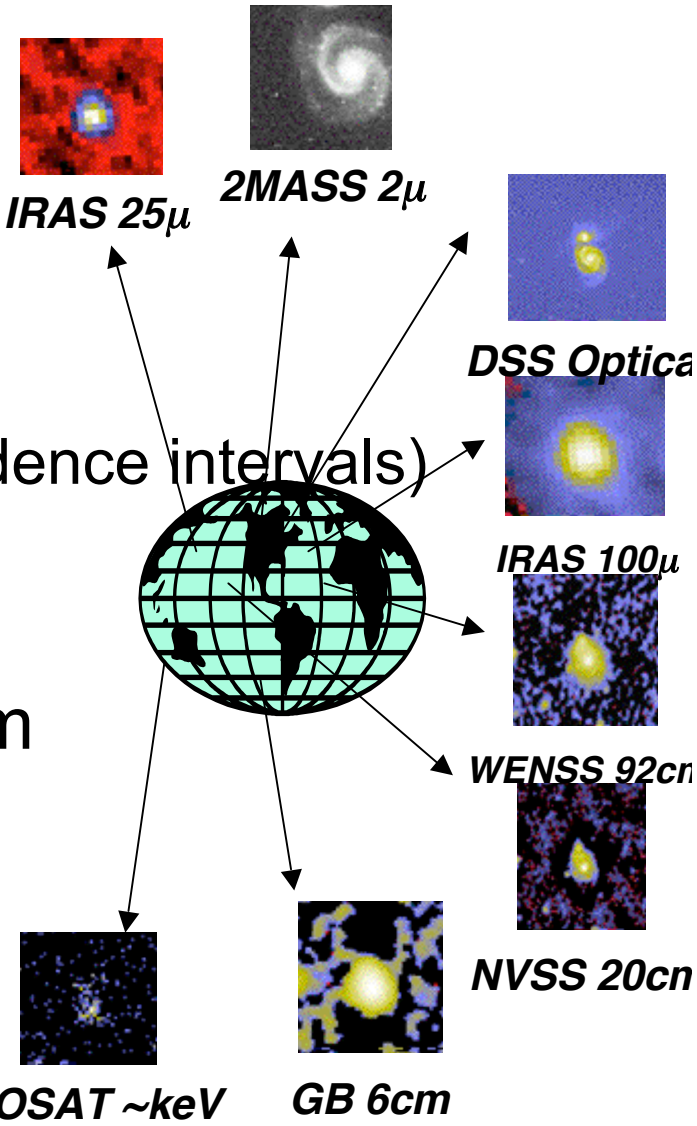
http://www.us-vo.org/ http://www.ivoa.net/

Jim Gray, "Online Science the New Computational Science," Microsoft Research.

PENNSTATE ARL

Why are Computer Scientists interested in Astronomy Data?

- It has no commercial value
  - –No privacy concerns
  - –Can freely share results with others
  - –Great for experimenting with algorithms
- It is real and well documented
  - –**High-dimensional data** (with confidence intervals)
  - –**Spatial** data
  - –**Temporal** data
- Many **different instruments** from many **different places** and many **different times**
- Federation is a goal
- There is a lot of it (petabytes)

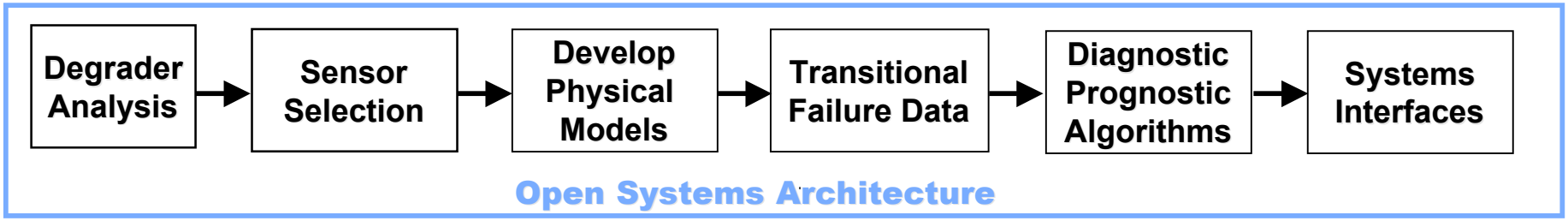


Jim Gray, "Online Science the New Computational Science," Microsoft Research.

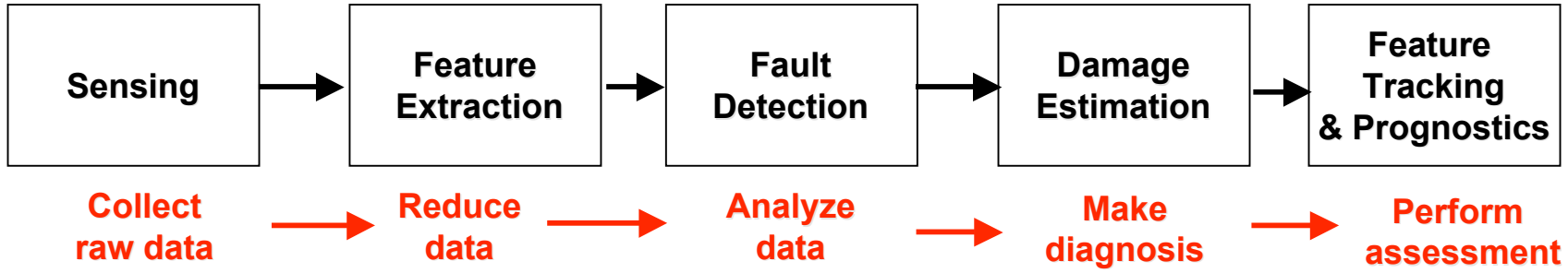
ISHM Development and Implementation

Where can we apply data mining techniques in ISHM development and implementation?

Development

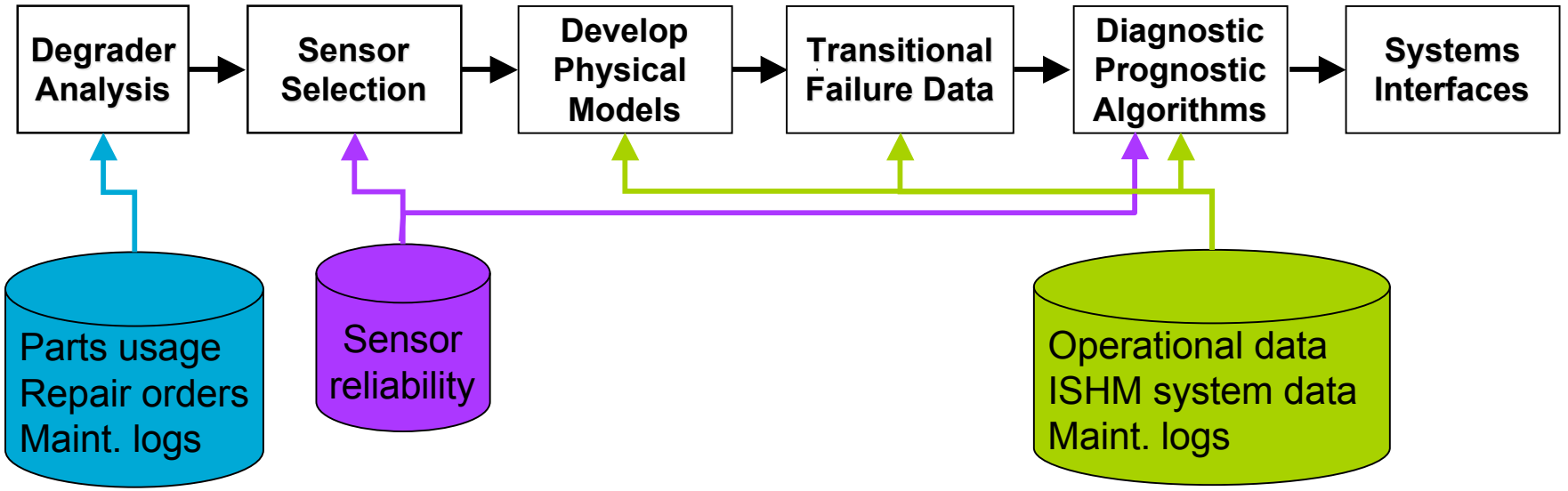


Implementation



Data Mining in ISHM Development

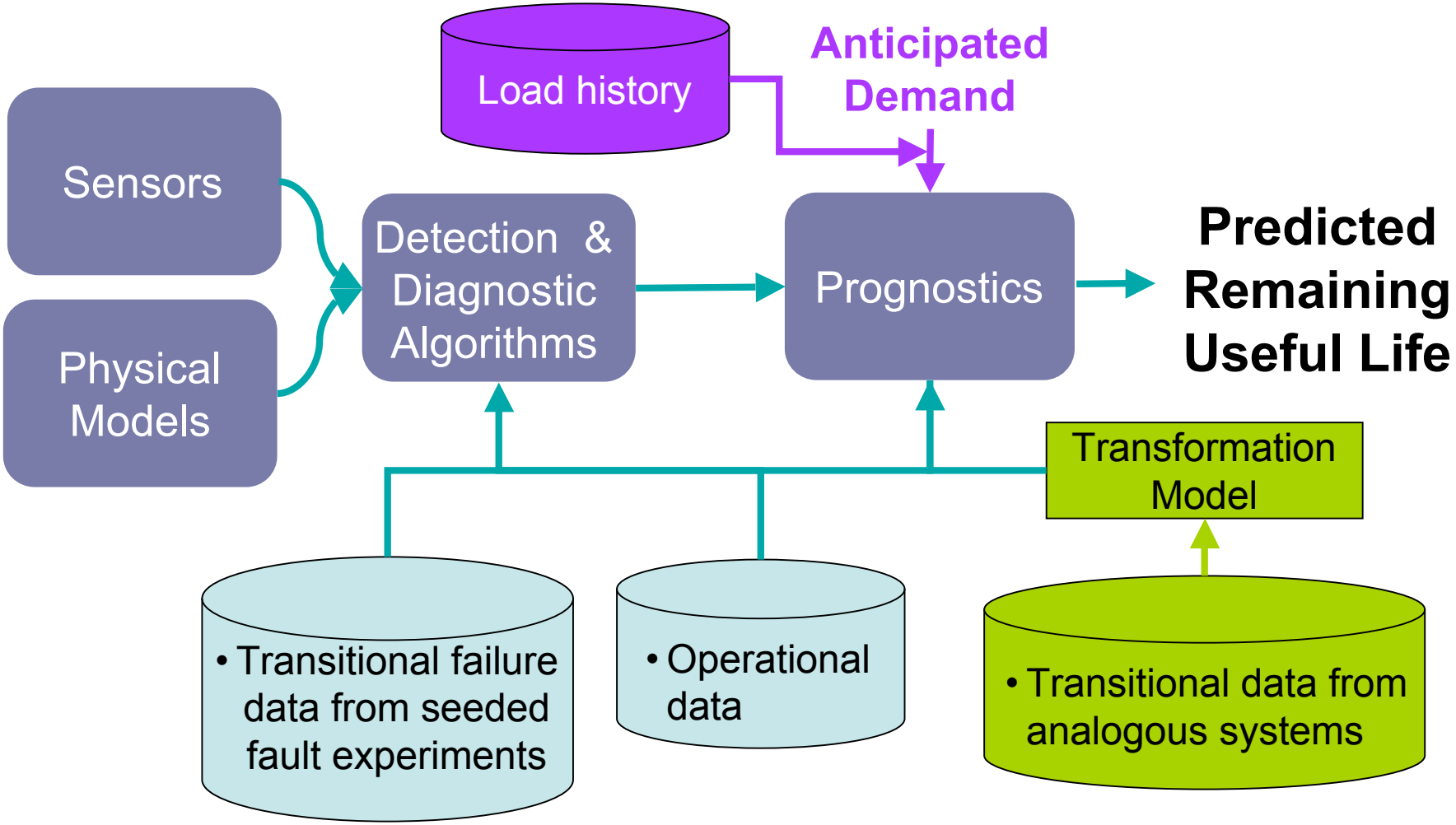
- Initial design and development
- ISHM system verification and validation
- Continuous ISHM system improvement (borrowing techniques from manufacturing and quality control)



Biggest payoff may be in collection and analysis of transitional failure data across large numbers systems

PENNSTATE ARL

Data Mining in Prognostics

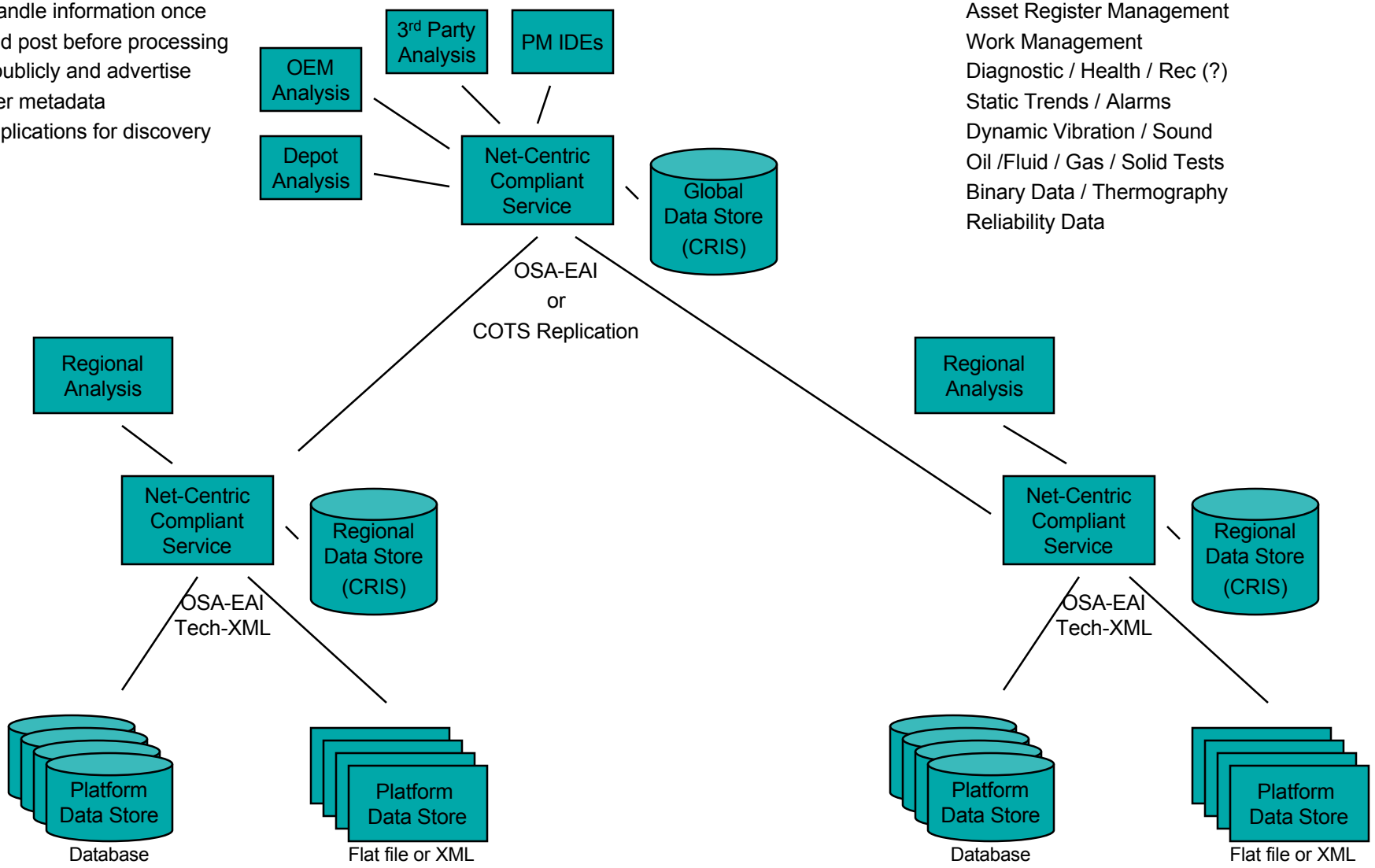


Data mining may allow access to much larger sets of transitional data for assessing fault severity and predicting remaining useful life

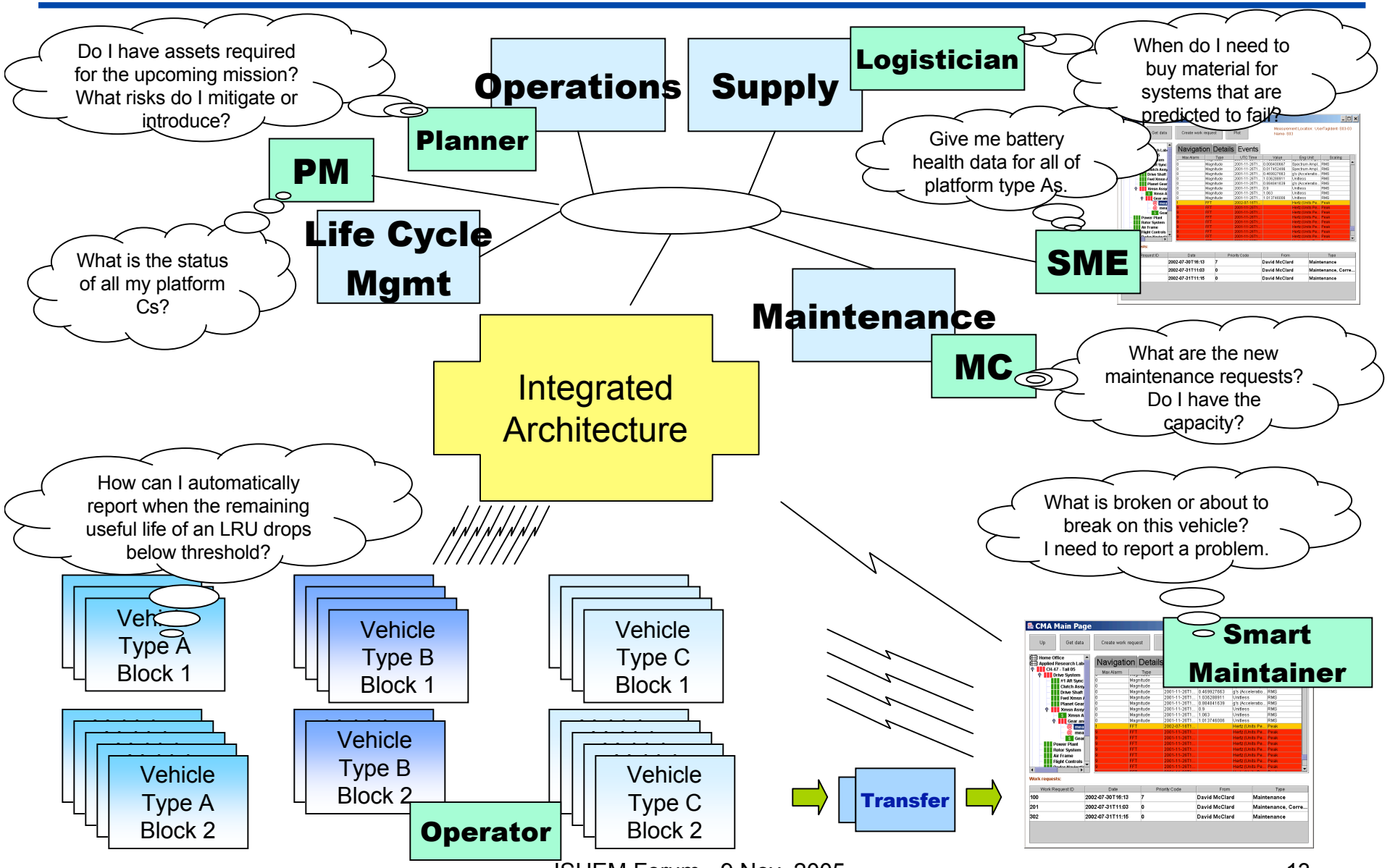
Managing System Health Information

Net Centric Attributes
Only handle information once
Tag and post before processing
Store publicly and advertise
Register metadata
Tag applications for discovery

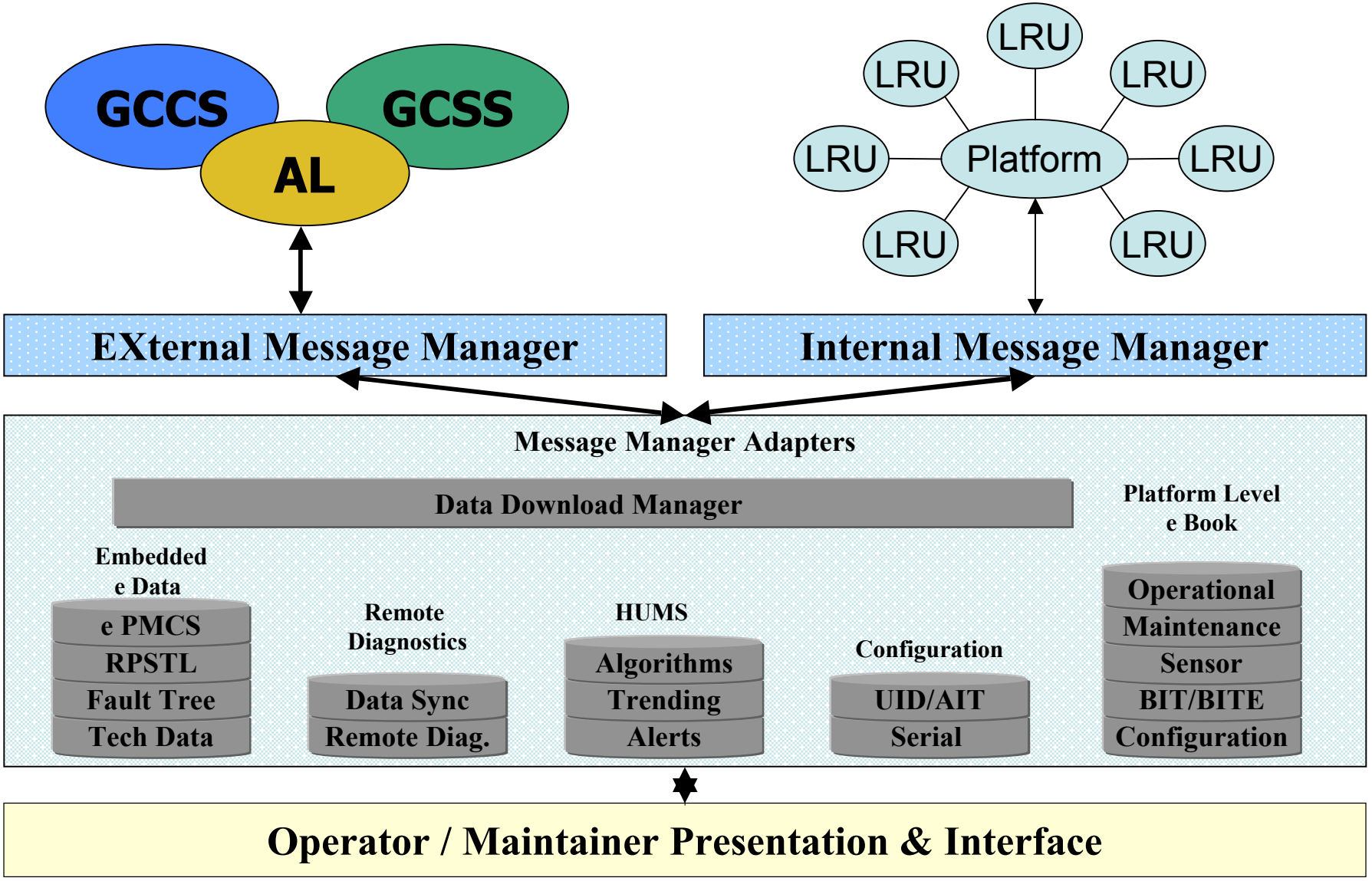
What types of data are we talking about?
Asset Register Management
Work Management
Diagnostic / Health / Rec (?)
Static Trends / Alarms
Dynamic Vibration / Sound
Oil /Fluid / Gas / Solid Tests
Binary Data / Thermography
Reliability Data



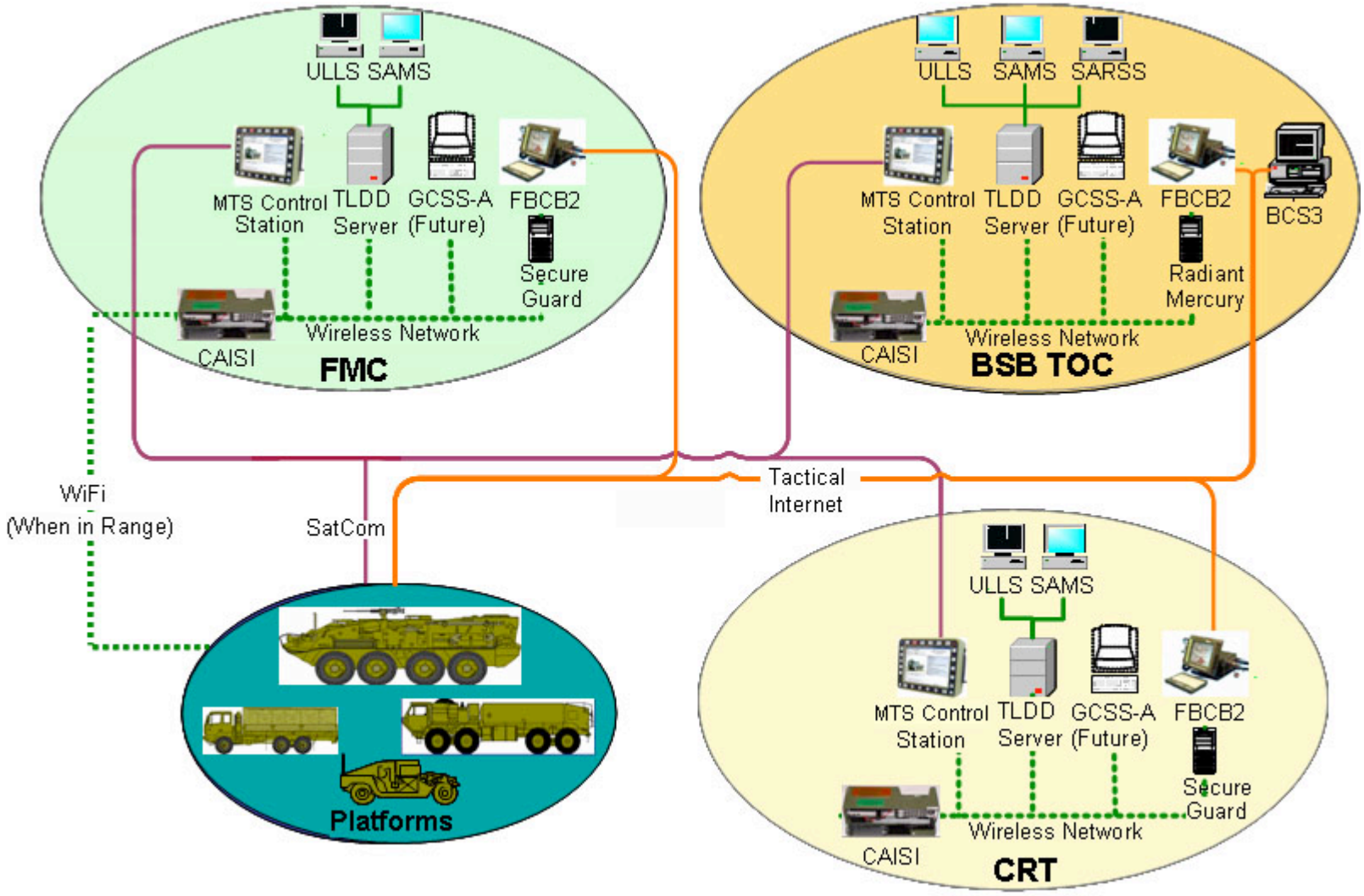
Leveraging Platform Asset Health Information



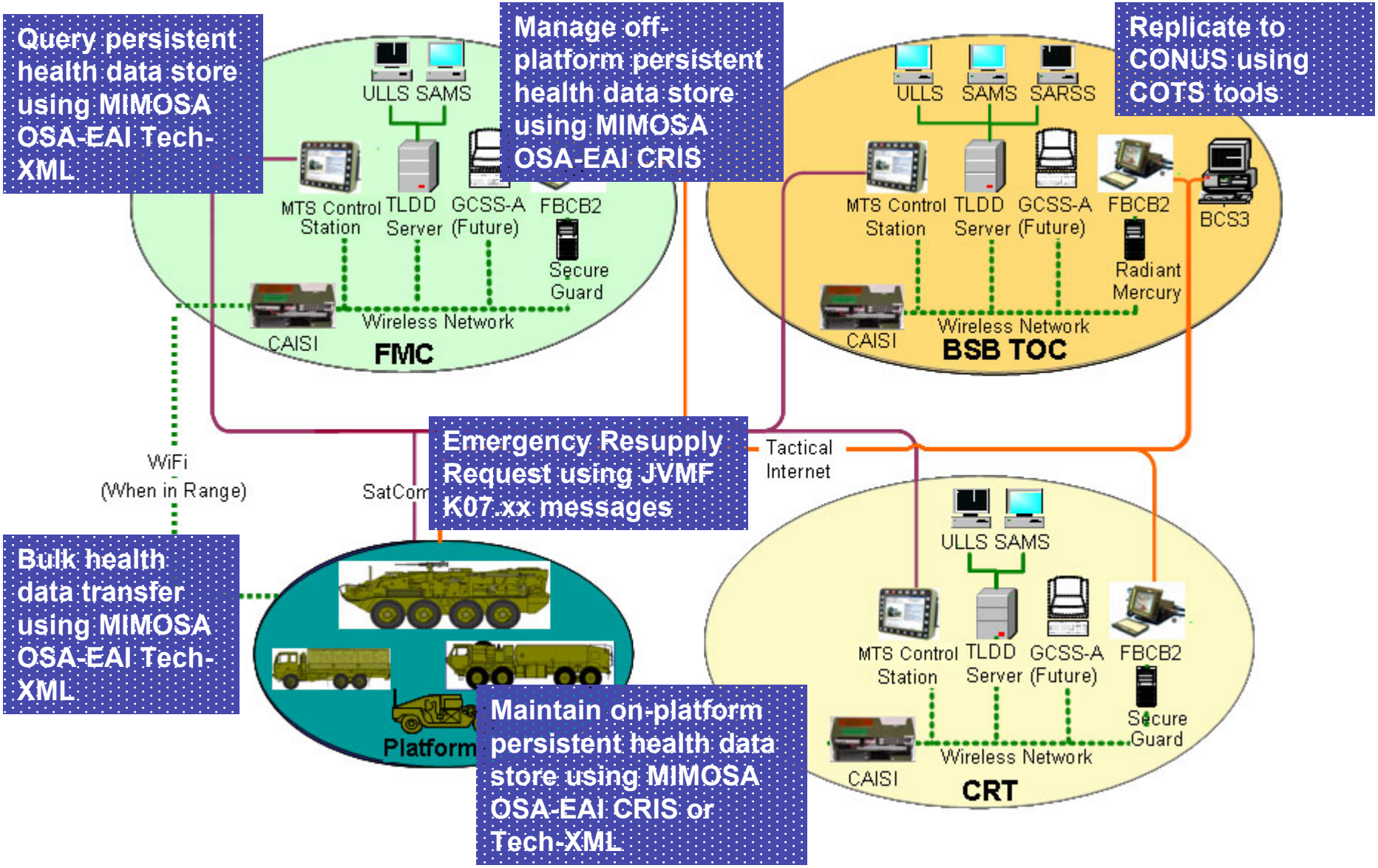
USMC Autonomic Logistics Architecture



US Army Common Logistics Operating Environment

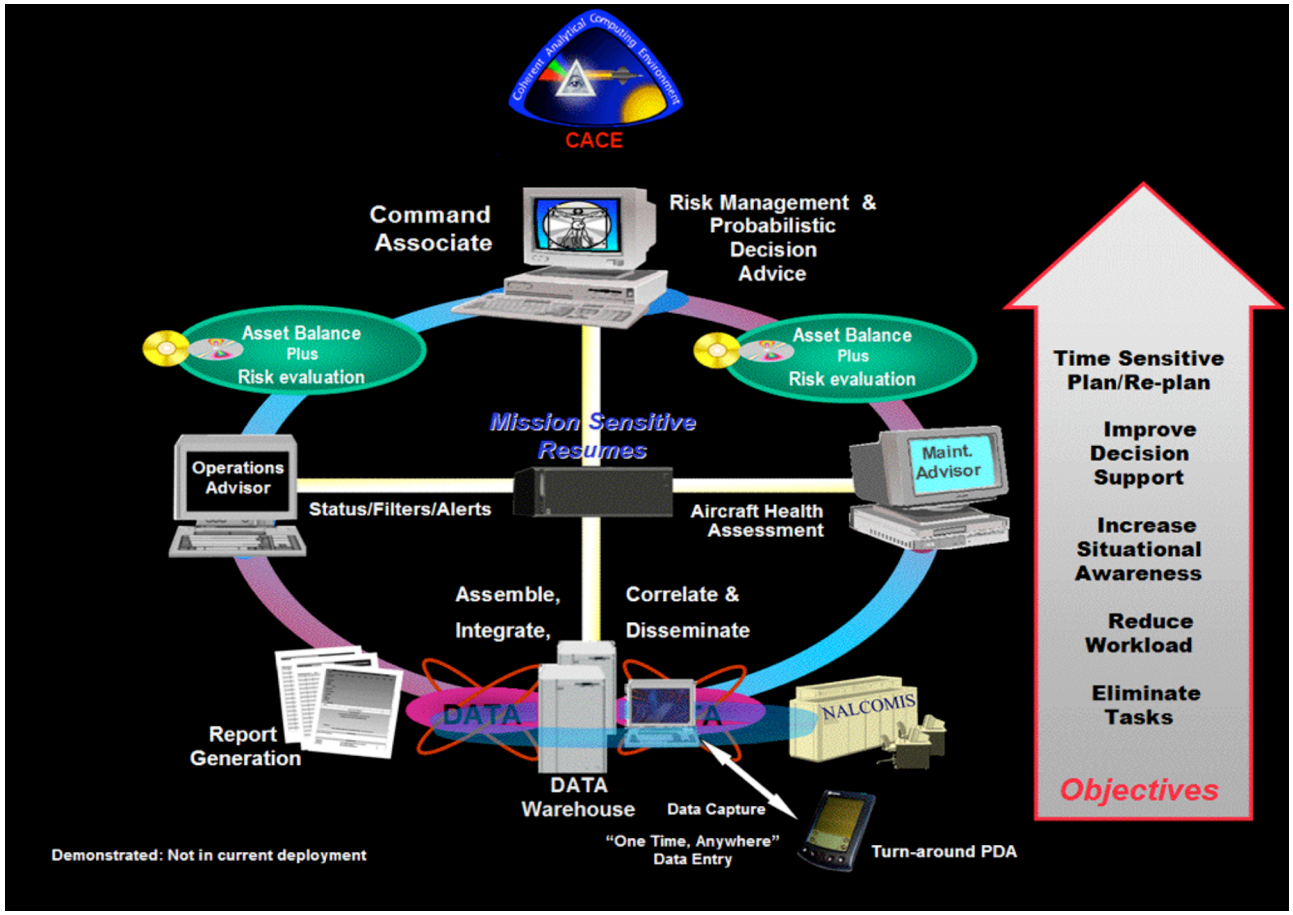


US Army Common Logistics Operating Environment



CACE - Coherent Analytical Computing Environment

CACE is an example of a data mining application which taps into databases of maintenance requests, repair schedules, and training schedules to optimize flight operations

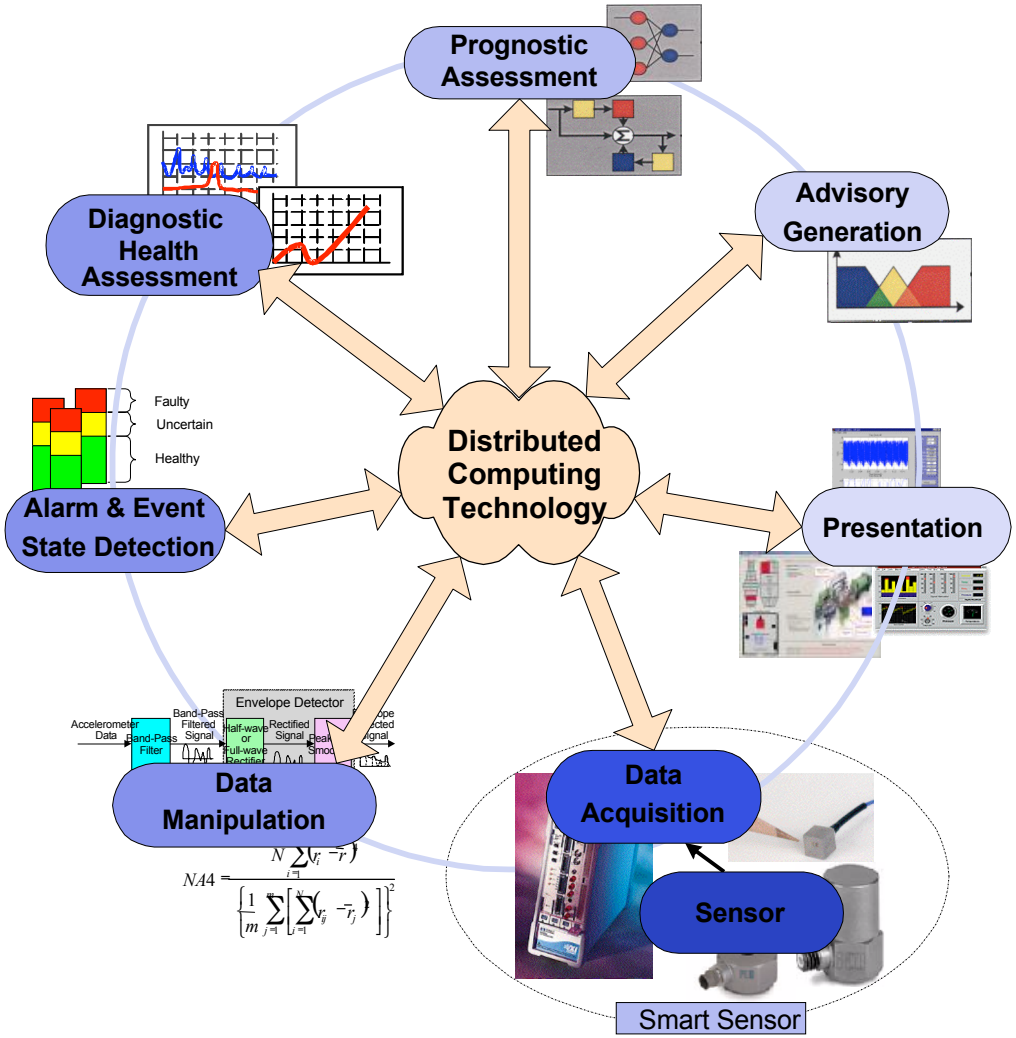


PENNSTATE ARL

Access to Data

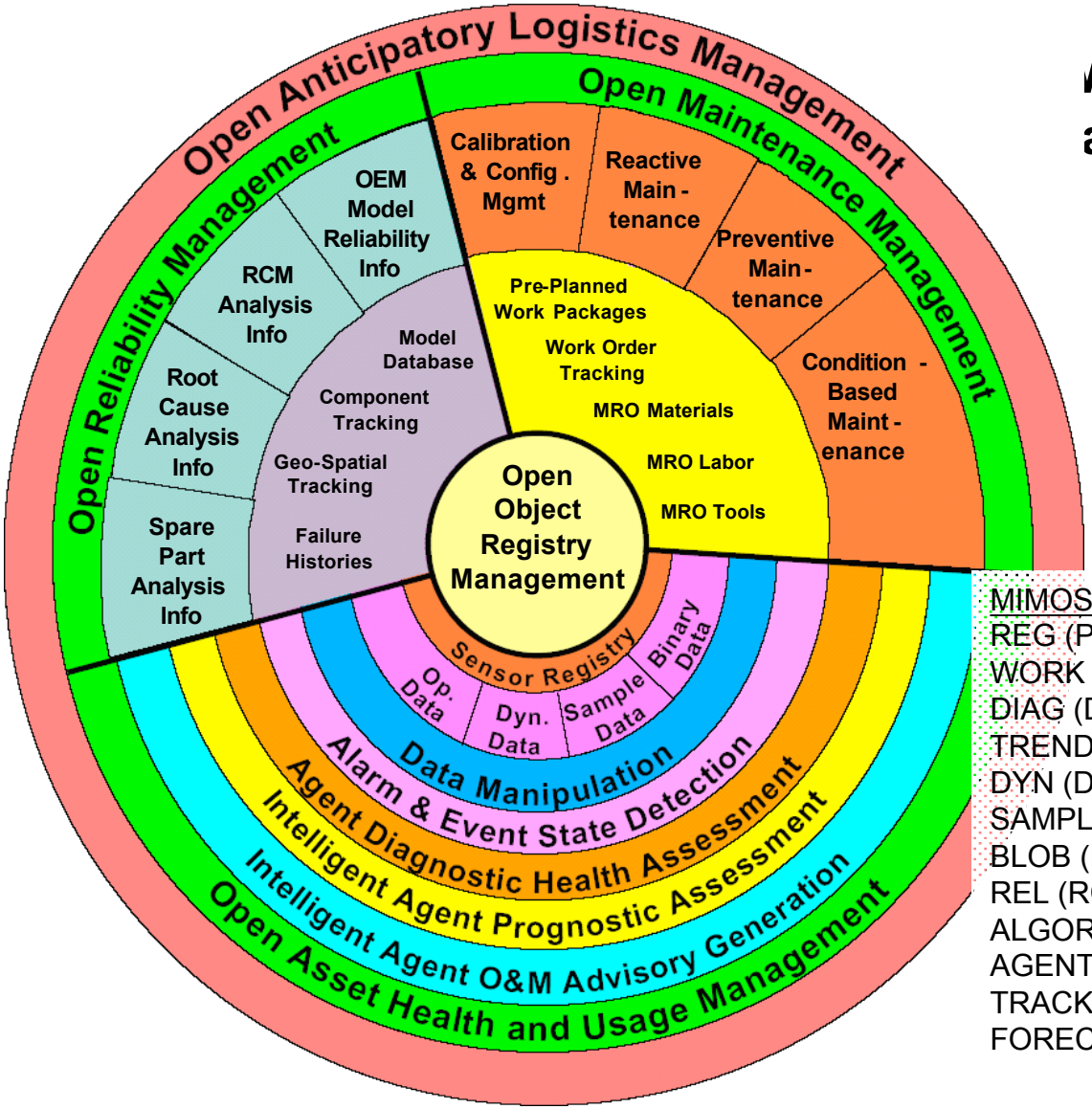
- Need access to data for all information customers
  - Embedded ISHM systems
  - Operators
  - Maintainers
  - Planners
  - Program managers
  - Design Engineers
- Latency of information will determine value to different customers
- Who is responsible for maintaining the data warehouse?

Open Systems Architecture for Condition Based Maintenance



- Standardized architecture for health and condition monitoring systems
- Breaks monitoring system into functional layers
- Defines inputs and outputs required for each layer
- Modules not confined to one locale
- Middleware independent

Standards for Enabling System Management

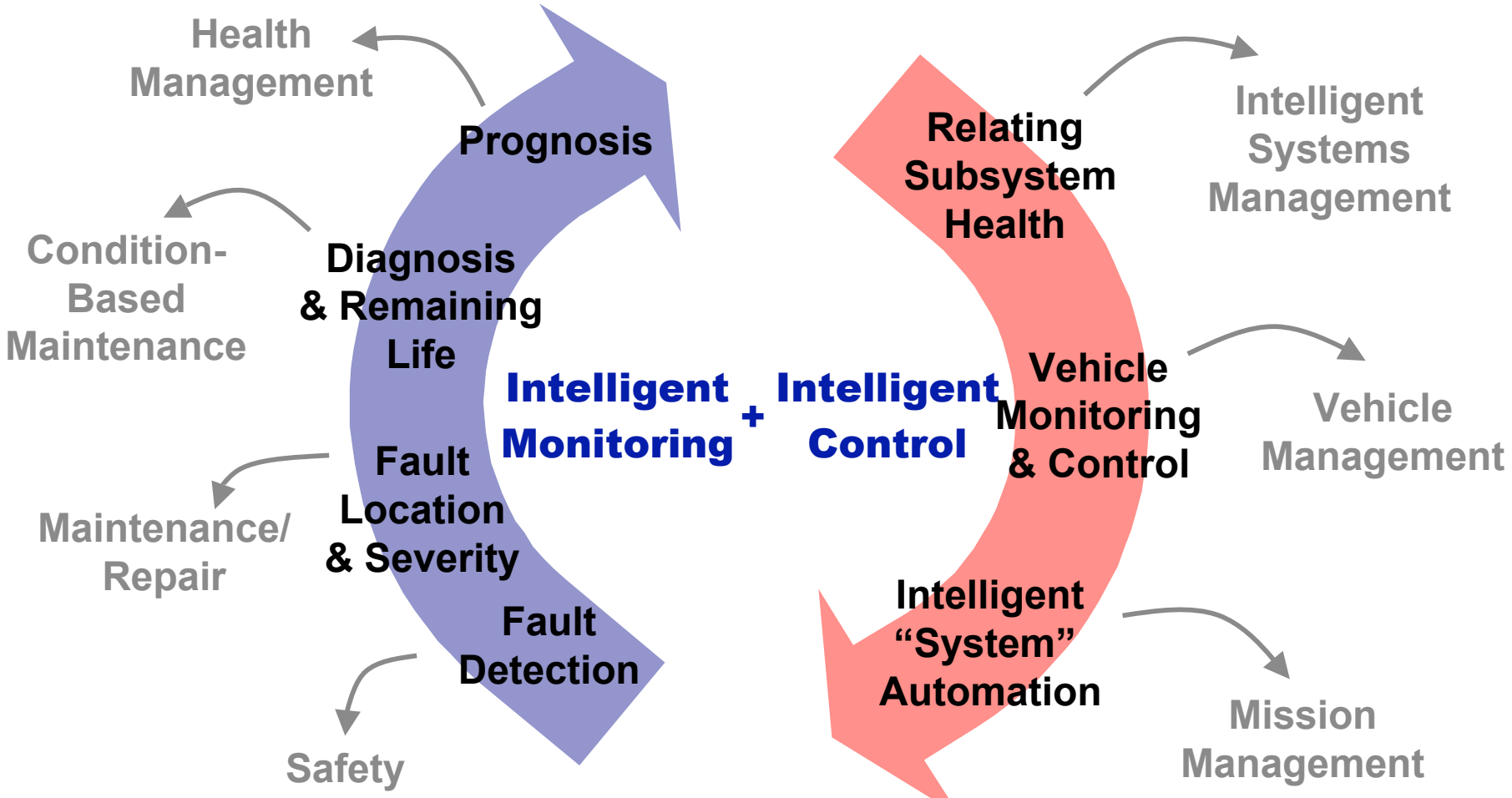


Machinery Information Management Open Standards Alliance (MIMOSA)

MIMOSA Technology Types
REG (Physical Asset Register Management)
WORK (O&M Agent Work Management)
DIAG (Diagnostics / Prognostics / Health Assessment)
TREND (Operational Scalar Data & Alarms)
DYN (Dynamic Vibration/Sound Data & Alarms)
SAMPLE (Oil/Fluid/Gas/Solid Test Data & Alarms)
BLOB (Binary Data/Thermography Data & Alarms)
REL (RCM/FMECA/Model Reliability Information)
ALGORITHM (Algorithm Management Information)
AGENT (Intelligent Agent Management Information)
TRACK (Physical Asset GeoSpatial Tracking Info.)
FORECAST (Capability Forecasting & Projections)

PENNSTATE ARL

Beyond System Health Monitoring



Need techniques to capture decisions and actions in addition to data and permit machines to mine those high level control "lessons learned"

PENNSTATE ARL

Other Data Mining Challenges

- Data management
- Data storage
- Indexing data for efficient computer access
- Ownership of data, data rights, and security
- Distributed processing architectures
- Data summarization, trend detection anomaly detection are key technologies
- Translation models

PENNSTATE ARL

Conclusions

- Data mining is the search for hidden relationships and meaning in data
- Applications in ISHM include modeling, diagnostics, and prognostics
- Useful for system development and for closed-loop improvement of systems
- Biggest payoff could be in creation of virtual transitional failure data sets
- Information customers are driving for network-centric applications so databases and data warehouses will be there – we need to be sure the data is useful for ISHM